SUPPLEMENTAL MATERIAL

TGF-β/Smad signaling through DOCK4 facilitates lung adenocarcinoma metastasis

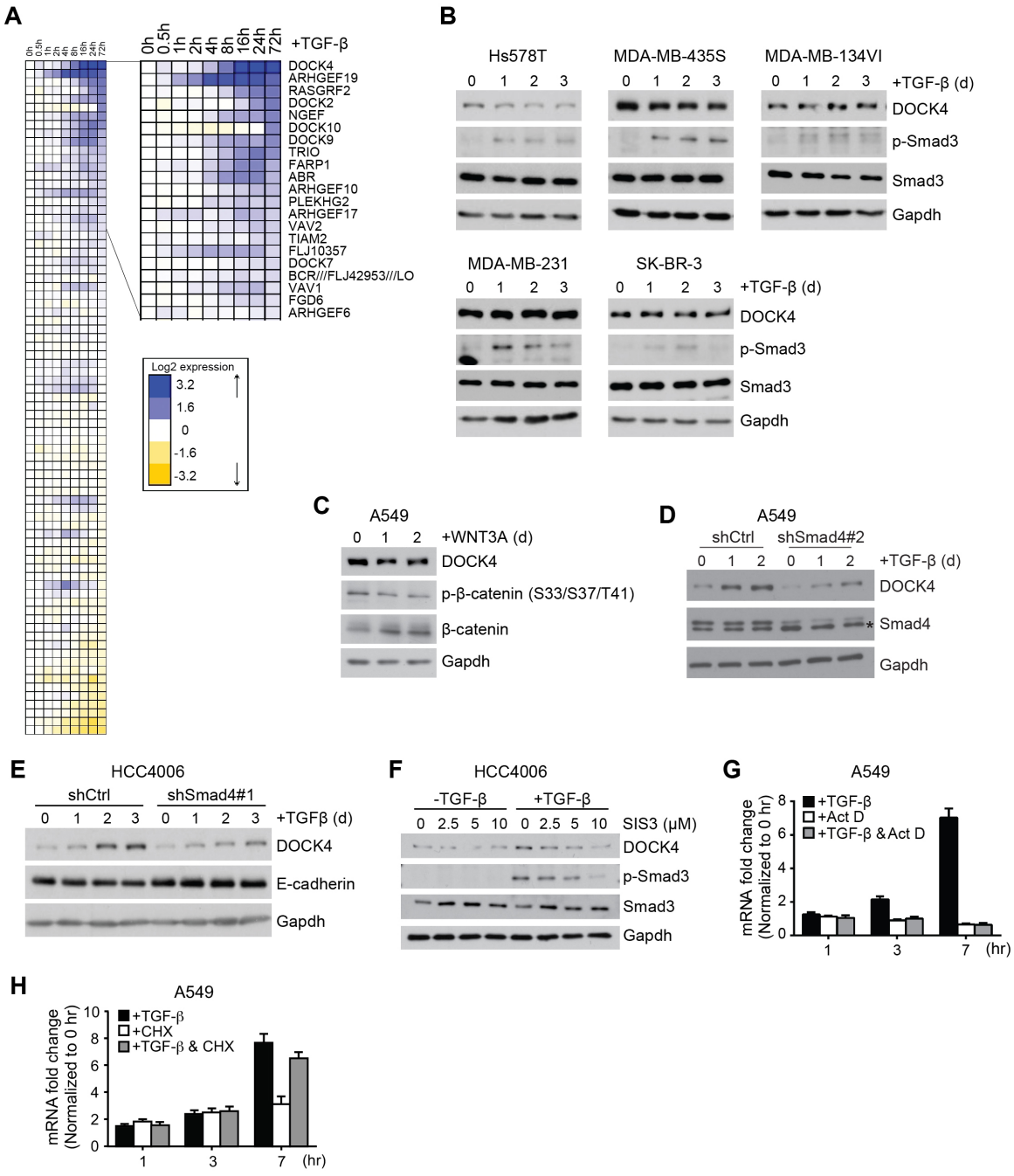
Jia-Ray Yu, Yilin Tai, Ying Jin, Molly C. Hammell, J. Erby Wilkinson, Jae-Seok Roe, Christopher R. Vakoc, and Linda Van Aelst

Ten Supplemental Figures

Four Supplemental Tables

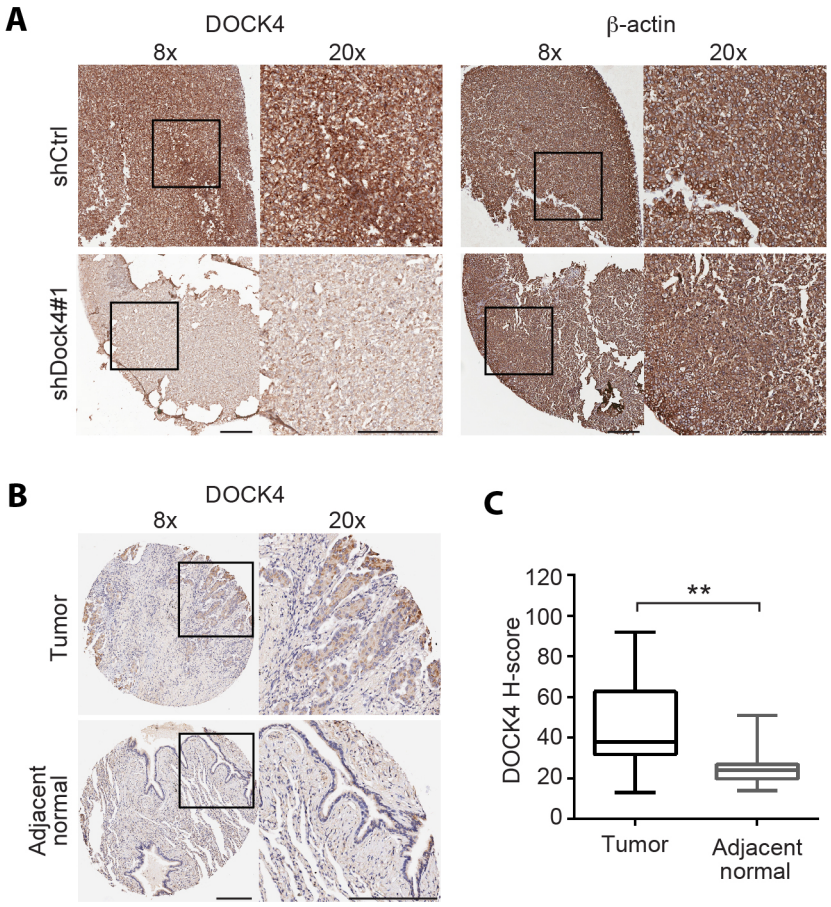
Supplemental Materials and Methods

Supplemental References

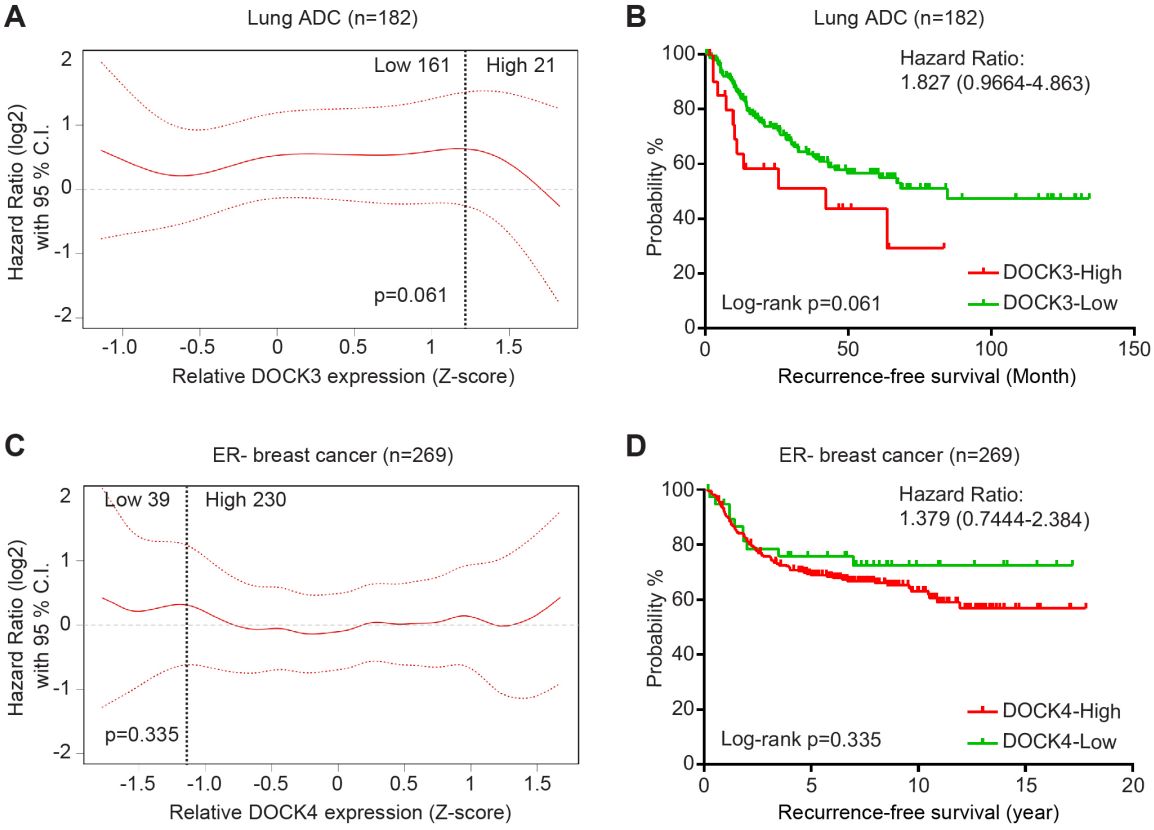


Supplemental Figure 1, related to Figure 1. TGF-β/Smad signaling induces DOCK4 expression in lung ADC cells, but not breast cancer cells. *(A)* Heat map showing differential expression of all 83 Rho family GEFs in TGF-β treated A549 cells over a 72-h time window. Gene expression data are presented on a log2 scale. Blue and yellow indicate upregulation and downregulation, respectively, by TGF-β. The top 21 TGF-β upregulated Rho family-GEFs are listed on the right. Original data were retrieved from Gene Expression Omnibus (GEO) with accession number GSE17708. (*B*) Western blot analysis of DOCK4, p-Smad3, and Smad3 in a panel of breast cancer cell lines treated

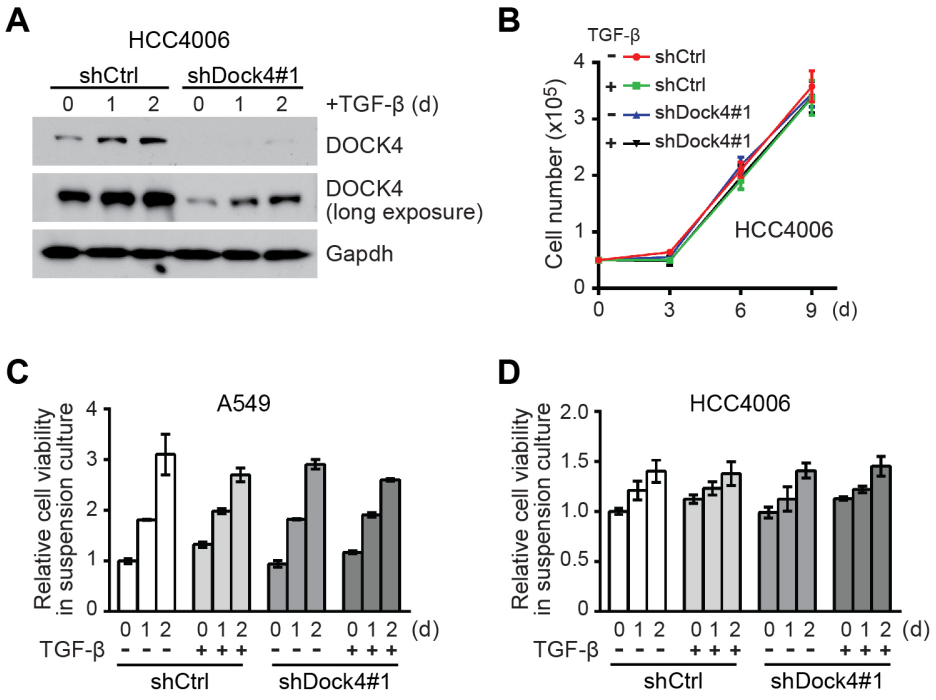
with 2 ng/ml TGF-β over a 3-d time window. Gapdh was used as a loading control. (*C*) Western blot analysis of DOCK4, p-β-catenin, and β-catenin in A549 cells treated with 100 ng/ml WNT3A over a 2-d time window. (*D*) Western blot analysis of DOCK4 and Smad4 in shCtrl- and shSmad4#2-expressing A549 cells treated with 2 ng/ml TGF-β over a 2-d time window. Asterisk depicts a non-specific band. (*E*) Western blot analysis of DOCK4 and E-cadherin in shCtrl- and shSmad4#1-expressing HCC4006 cells treated with 2 ng/ml TGF-β over a 3-d time window. (*F*) Western blot analysis of DOCK4, p-Smad3, and Smad3 in HCC4006 cells treated with 0 to 10 μM SIS3 and/or 2 ng/ml TGF-β for 24 h. (*G*) qPCR analysis of DOCK4 mRNA in A549 cells treated with 2 ng/ml TGF-β, 2 μg/ml Actinomycin D (Act D), or both. (*H*) qPCR analysis of DOCK4 mRNA in A549 cells treated with 2 ng/ml TGF-β, 50 μg/ml Cycloheximide (CHX), or both. qPCR data in G and H were normalized to Gapdh, and presented as mean ± SD (n = 3).



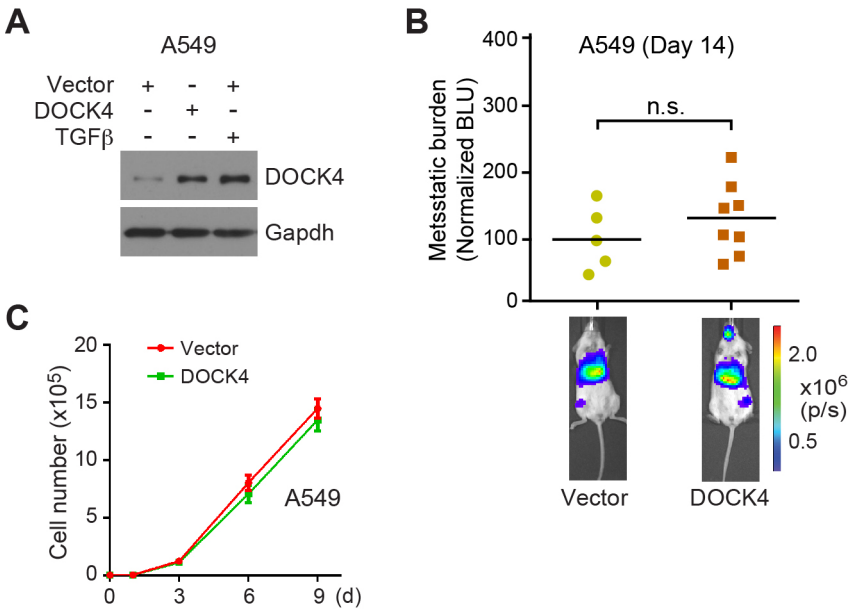
Supplemental Figure 2, related to Figure 2. Validation of anti-DOCK4 antibody for immunohistochemical (IHC) staining, and DOCK4 expression in primary lung ADC and tumor adjacent normal tissues. *(A)* shCtrl- and shDock4#1-expressing A549 cell pellets processed with acidic formalin fixation and paraffin embedding, and stained with antibodies against DOCK4 (left) and β-actin (right). Boxed regions are enlarged and shown on the right. Scale bars, 250 μm. *(B)* Representative images of DOCK4 IHC staining of a tissue microarray containing human lung ADC tissues and tumor adjacent normal tissues. Nuclei were counterstained with hematoxylin. Boxed regions are enlarged and shown on the right. Scale bars, 250 μm. *(C)* H-scores of DOCK4 expression in 16 pairs of matched human lung ADC tissues and their adjacent normal tissues that contain sufficient epithelium area for IHC quantification. $^{**}P < 0.01$ by a paired Student's *t*-test.



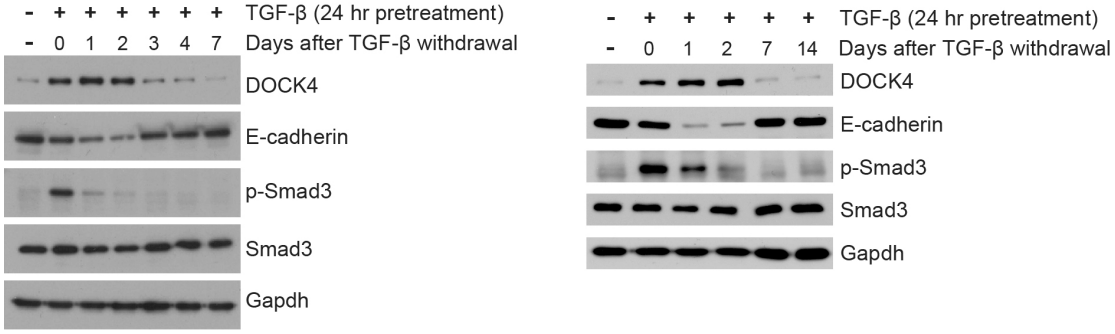
Supplemental Figure 3, related to Figure 2. DOCK3 and DOCK4 are not significantly associated with recurrence-free survival in lung ADC and in estrogen-receptor (ER)-negative breast cancer patients, respectively. *(A)* Hazard ratio plot in function of DOCK3 expression based on gene expression and recurrence-free survival data for a cohort of 182 lung ADC patients. *(B)* Kaplan-Meier survival curve for low and high DOCK3 expression groups indicated in *(A)*. *(C)* Hazard ratio plot in function of DOCK4 expression based on gene expression and recurrence-free survival data for a cohort of 269 ER-negative (ER-) breast cancer patients. Of note, publicly available breast cancer datasets were filtered to include only patients with ER-negative breast cancer, as TGF-β activity was found to be correlated with distant metastasis in ER-negative, but not ER-positive, breast tumors (Padua et al. 2008). *(D)* Kaplan-Meier survival curve for low and high DOCK4 expression groups indicated in *(C)*. Dotted lines in A and C indicate the cut-offs at the points for the highest hazard ratio with confidence interval 95% to define low/high expression groups.



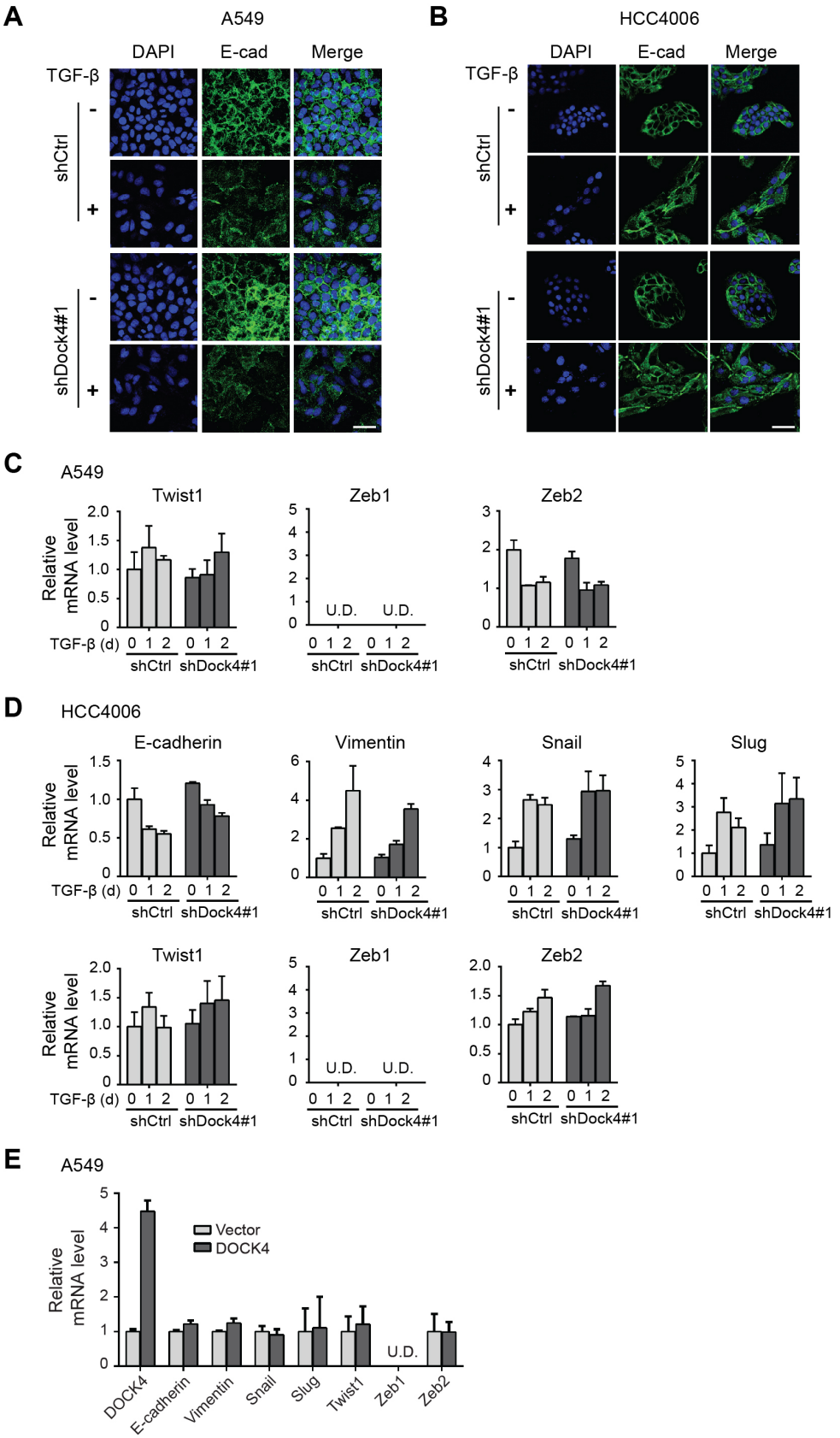
Supplemental Figure 4, related to Figure 3. DOCK4 is dispensable for lung ADC cell proliferation in 2D culture and viability in suspension culture. (*A*) Western blot analysis of DOCK4 in shCtrl- and shDock4#1-expressing HCC4006 cells treated with 2 ng/ml TGF-β over a 2-d time window. Gapdh was used as a loading control. (*B*) Growth curve of TGF-β pretreated (24 h, +) and untreated (-) shCtrl- and shDock4#1-expressing HCC4006 cells over a 9-d time window. (*C*,*D*) MTT cell viability assay of TGF-β pretreated (24 h, +) and untreated (-) shCtrl- and shDock4#1-expressing A549 (*C*) and HCC4006 (*D*) cells in suspension culture over a 2-d time window. Data in B, C and D represent mean ± SD ($n = 3$).



Supplemental Figure 5, related to Figure 3. Ectopic expression of DOCK4 does not alter the metastatic potential or growth properties of lung ADC cells. (*A*) Western blot analysis of DOCK4 expression in vector- and DOCK4-expressing A549 cells, and in vector-expressing A459 cells treated with 2 ng/ml TGF-β for 24 h. (*B*) Analysis of metastatic burden in NSG mice intracardially injected with vector- or DOCK4-expressing A549 cells. Top, dot plots of metastatic burden at day 14. BLU signals were normalized to day 0, and then to a value of 100 for vector control condition. Bottom, representative images of mice with metastases. n = 5 (vector) and 8 (DOCK4) mice. P value was calculated by an unpaired Mann-Whitney test. n.s., $P \geq 0.05$. (*C*) Growth curve of A549 cells stably expressing vector or DOCK4 over a 9-d time window. Data represent mean ± SD (n = 3).

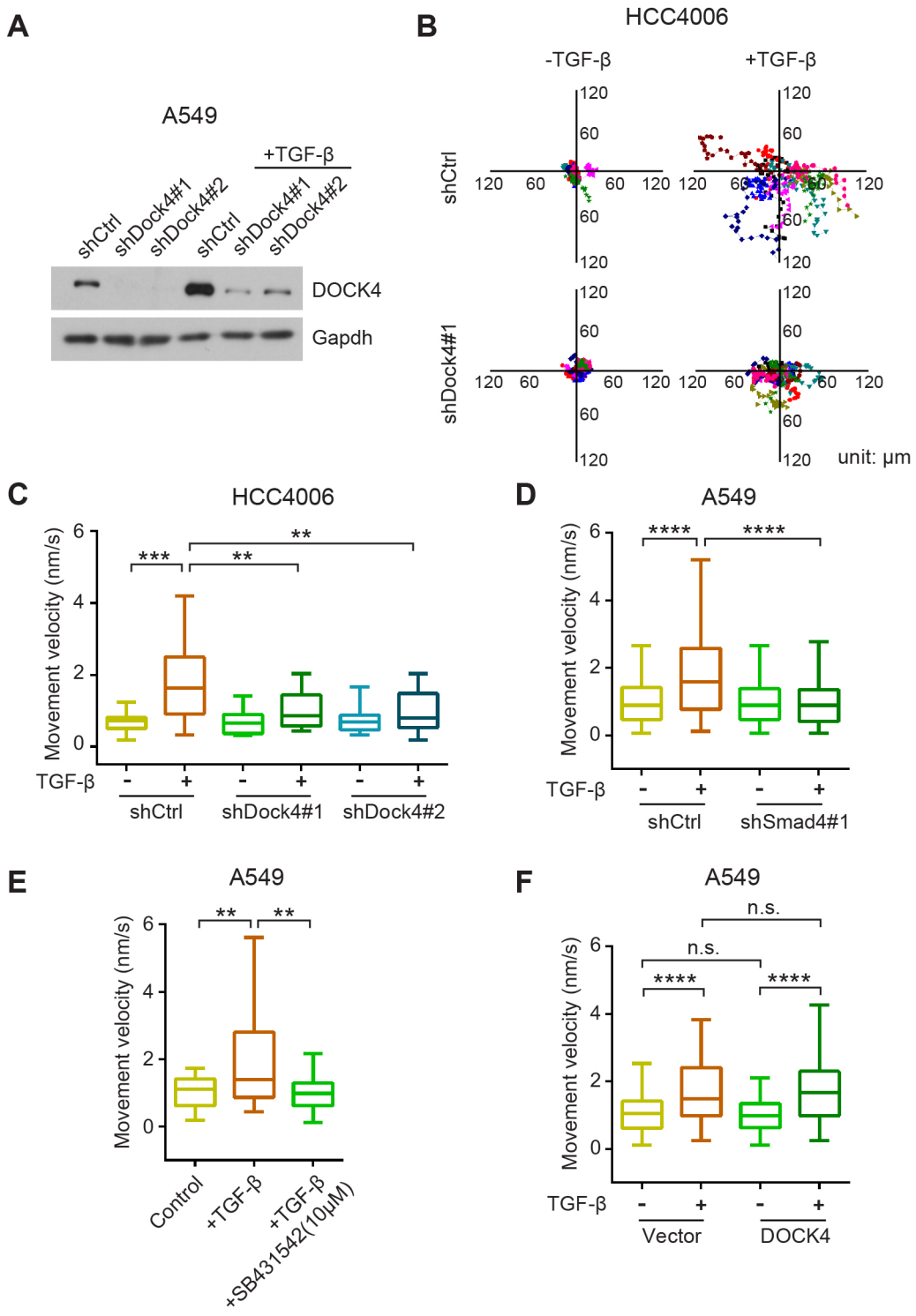


Supplemental Figure 6, related to Figure 4. Downregulation of TGF-β signaling activity, and reversal of E-cadherin and DOCK4 expression following the removal of TGF-β. Western blot analyses of DOCK4, E-cadherin, p-Smad3, and Smad3 in A549 cells treated with TGF-β for 24 h followed by TGF-β withdrawal at indicated time points over a 7-d (left) and 14-d time window (right). Gapdh was used as a loading control.



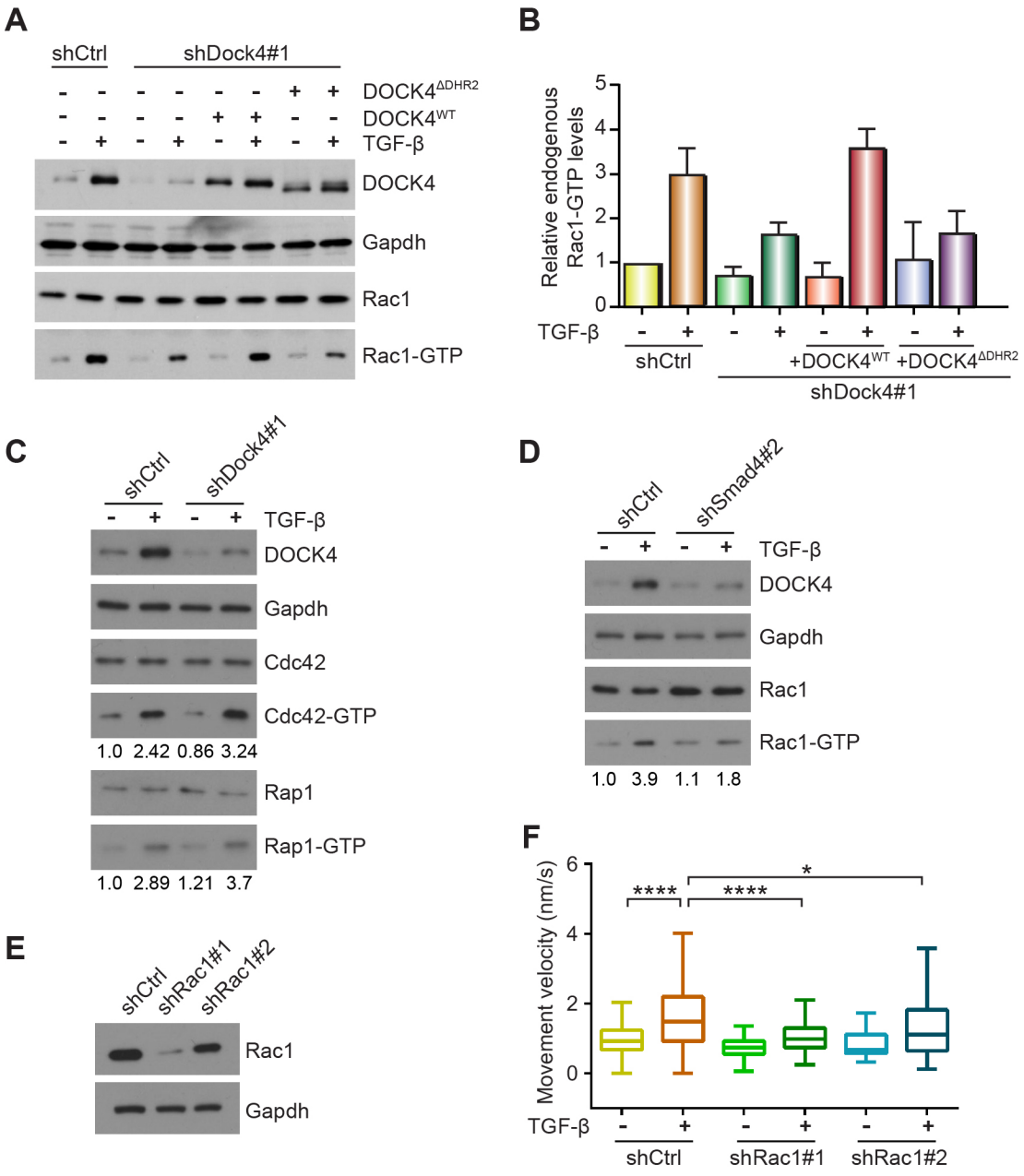
Supplemental Figure 7, related to Figure 5. DOCK4 is dispensable for TGF-β induced EMT. (*A*,*B*) Representative confocal images of TGF-β treated (24 h, +) and untreated (-)

shCtrl- and shDock4#1-expressing A549 (*A*) and HCC4006 (*B*) cells immunostained for E-cadherin (green). Nuclei were counterstained with DAPI (blue). Scale bars, 50 μm. (*C*, *D*) qPCR analysis of indicated EMT markers mRNAs in shCtrl- and shDock4#1-expressing A549 (*C*) and HCC4006 (*D*) cells treated with TGF-β over a 2-d time window. Data represent mean ± SD (n = 3). U.D, undetectable. (*E*) qPCR analysis of indicated EMT markers mRNAs in vector- and DOCK4-expressing A549 cells. Data represent mean ± SD (n = 3). U.D, undetectable.



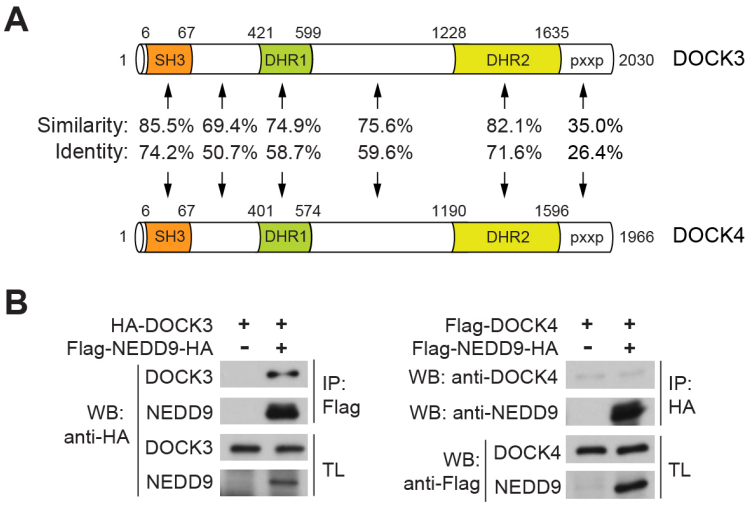
Supplemental Figure 8, related to Figure 5. DOCK4 and the Smad signaling pathway mediate TGF-β-induced single cell motility, and ectopic expression of DOCK4 alone does not enhance single cell motility. (*A*) Western blot analysis of endogenous DOCK4 expression in A549 cells expressing shCtrl, shDock4#1, or shDock4#2 left untreated or treated with 2 ng/ml TGF-β for 24 h. Gapdh was used as a loading control. (*B*)

Representative movement trajectories of single shCtrl- or shDock4#1-expressing HCC4006 cells, left untreated or pretreated (24 h) with 2 ng/ml TGF-β, obtained by live-cell imaging over a 4.5-h time window. Each trajectory represents the movement of a single cell, and individual dots designate a frame of 6 min. (*C*) Quantification of velocity of movement of shCtrl-, shDock4#1- or shDock4#2-expressing HCC4006 cells in *B*. n = 14-26 cells/condition. (*D*) Quantification of velocity of movement of shCtrl- or shSmad4#1-expressing A549 cells, left untreated or pretreated (24 h) with 2 ng/ml TGF-β. n = 82-152 cells/condition. (*E*) Quantification of velocity of movement of A549 cells left untreated, pretreated (24 h) with 2 ng/ml TGF-β, or pretreated (24 h) with 2 ng/ml TGF-β and 10 μM SB431542. n = 31-72 cells/condition. (*F*) Quantification of velocity of movement of vector- or DOCK4-expressing A549 cells left untreated or pretreated (24 h) with 2 ng/ml TGF-β. n = 64-129 cells/condition. Data in C, D, E, and F are presented as Tukey box plots, representing the upper quartile + 1.5 interquartile range (IQR; upper end of whisker), upper quartile (top of box), mean (band in box), lower quartile (bottom of box), and lower quartile - 1.5 IQR (lower end of whisker). $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, n.s., $P \geq 0.05$ by an unpaired Mann-Whitney test.



Supplemental Figure 9, related to Figure 5. DOCK4 mediates TGF-β-induced Rac1, but not Rap1 or Cdc42, activation. (*A*) Activation of Rac1 measured with p21-activated kinase Rac/Cdc42-binding domain (PBD) pull-down assays. A549 cells expressing shCtrl, shDock4#1, shDock4#1 + DOCK4^{WT}, or shDock4#1 + DOCK4$^{\Delta DHR2}$ were treated with TGF-β (24 h, +) or left untreated (-), and GTP-bound Rac1 was precipitated from detergent extracts with GST-PBD. GST-PBD-bound Rac1-GTP, total Rac1, and DOCK4 levels in cell lysates were detected by immunoblotting with anti-Rac1 and anti-DOCK4 antibodies. Gapdh was included as a loading control. (*B*) Quantification of the ratio of Rac1-GTP versus total Rac1 from 3 independent experiments. Data were normalized to control condition (shCtrl, TGF-β -) and presented as mean ± SD. (*C*) Activation of Rap1 and Cdc42 measured with RalGDS-Rap-binding domain (RBD) and PBD pull-down

assays, respectively. A549 cells expressing shCtrl or shDock4#1 were treated with TGF-β (24 h, +) or left untreated (-), and GTP-bound Rap1 or Cdc42 was precipitated from detergent extracts with GST-RalGDS-RBD or GST-PBD, respectively. GST-RalGDS-RBD-bound Rap1-GTP, total Rap1, GST-PBD-bound Cdc42-GTP, total Cdc42, and DOCK4 levels in cell lysates were detected by immunoblotting with anti-Rap1, anti-Cdc42 and anti-DOCK4 antibodies. The relative amount of Cdc42-GTP and Rap1-GTP in extracts (compared to shCtrl, TGF-β -) is indicated in the bottom panels. (*D*) Activation of Rac1 measured as described in (*A*) in shCtrl- or shSmad4#2-expressing A549 cells left untreated or treated (24 h) with 2 ng/ml TGF-β. GST-PBD-bound Rac1-GTP, total Rac1, and DOCK4 levels in cell lysates were detected by immunoblotting with anti-Rac1 and anti-DOCK4 antibodies. The relative amount of Rac1-GTP in extracts (compared to shCtrl, TGF-β -) is indicated in the bottom panel. Data shown in *C* and *D* are representative of three independent experiments. (*E*) Western blot analysis of endogenous Rac1 protein in A549 cells expressing shCtrl, shRac1#1, or shRac1#2. (*F*) Quantification of velocity of movement of shCtrl-, shRac1#1-, or shRac1#2-expressing A549 cells left untreated or pretreated (24 h) with 2 ng/ml TGF-β. n = 48-77 cells/condition. Data are presented as Tukey box plots, representing the upper quartile + 1.5 interquartile range (IQR; upper end of whisker), upper quartile (top of box), mean (band in box), lower quartile (bottom of box), and lower quartile - 1.5 IQR (lower end of whisker). *$P < 0.05$, ****$P < 0.0001$ by an unpaired Mann-Whitney test.



Supplemental Figure 10, related to Figure 5 DOCK4 shares high protein sequence homology with DOCK3, but does not interact with NEDD9 (*A*) Protein sequence similarity and identity between DOCK3 and DOCK4 proteins. (*B*) Lysates from HEK-293T cells transiently expressing the indicated constructs were immunoprecipitated (IP) with an antibody to Flag (left panel) or HA (right panel) and analyzed by Western blot (WB) analysis with antibodies against HA (left panel) or DOCK4, NEDD9, or Flag (right panel). TL, total lysate.

Supplemental Table 1: List of cell lines used in this study. The identified mutations of common oncogenes/tumor suppressors are listed for each lung adenocarcinoma and breast cancer cell line.

Lung adenocarcinoma cell lines	**Genes**	**Protein sequence alteration**
A549	KRAS CDKN2A	G12S Deletion
H441	KRAS TP53	G12V R158L
HCC4006	EGFR	L747-E749 deletion, A750P
PC9	EGFR TP53	L746-A750 deletion R248Q
H1975	EGFR TP53 CDKN2A PIK3CA	L858R, T790M R273H E69* G118D
H1793	CDKN2A TP53	Deletion R209*; R273H

Breast cancer cell lines	**Genes**	**Protein sequence alteration**
MDA-MB-134-VI	TP53	E285K
MDA-MB-231	KRAS TP53 BRAF CDKN2A NF2	G13D R280K G464V Deletion E231*
MDA-MB-435S	TP53	G266E
SK-BR-3	TP53	R175H
Hs578T	HRAS TP53 CDKN2A PIK3R1	G12D V157F Deletion N453_T454insN

Supplemental Table 2: List of qPCR primers used in this study.

DOCK4	
Forward:	5'-GCATCTCTCGCTGGTTTGAAG-3'
Reverse:	5'-CAGGCACATAGTCAGGGGATT-3'
E-cadherin	
Forward:	5'-CGAGAGCTACACGTTCACGG-3'
Reverse:	5'-GGGTGTCGAGGGAAAAATAGG-3'
Vimentin	
Forward:	5'-GACGCCATCAACACCGAGTT-3'
Reverse:	5'-CTTTGTCGTTGGTTAGCTGGT-3'
Twist	
Forward:	5'-GTCCGCAGTCTTACGAGGAG-3'
Reverse:	5'-GCTTGAGGGTCTGAATCTTGCT-3'
Snail	
Forward:	5'-TCGGAAGCCTAACTACAGCGA-3'
Reverse:	5'-AGATGAGCATTGGCAGCGAG-3'
Slug	
Forward:	5'-CGAACTGGACACACATACAGTG-3'
Reverse:	5'-CTGAGGATCTCTGGTTGTGGT-3'
Zeb1	
Forward:	5'-GATGATGAATGCGAGTCAGATGC-3'
Reverse:	5'-ACAGCAGTGTCTTGTTGTTGT-3'
Zeb2	
Forward:	5'-CAAGAGGCGCAAACAAGCC-3'
Reverse:	5'-GGTTGGCAATACCGTCATCC-3'
Gapdh	
Forward:	5'-ACAACTTTGGTATCGTGGAAGG-3'
Reverse:	5'-GCCATCACGCCACAGTTTC-3'
DOCK4-SBE1	
Forward:	5'-GGAAGTGTAGCTTTCTATTAGG-3'
Reverse:	5'-GGCCAGACACATAGTAATGG-3'
DOCK4-SBE2	
Forward:	5'-CTGTGTGTCTCCTCCAAACC -3'
Reverse:	5'-CGAAACAGGAAGAAGGAACC-3'

Supplemental Table 3: Cox proportional univariate and multivariate analyses of recurrence free survival in a cohort of 182 lung ADC patients. (HR = hazard ratio; CI = confidence interval)

	Univariate analysis				Multivariate analysis			
Covariates	**HR**	**Low 95% CI**	**Up 95% CI**	**P-value**	**HR**	**Low 95% CI**	**Up 95% CI**	**P-value**
Age								
＜ 61	1							
≧ 61	1.508	0.911	2.495	0.11				
DOCK3 expression (Z-score)								
Low (＜1.15)	1							
High (≧1.15)	1.827	0.966	4.863	0.061				
DOCK4 expression (Z-score)								
Low (＜0)	1				1			
High (≧0)	2.074	1.295	3.297	**0.0024**	1.907	1.18	3.018	**0.0084**
Final stage								
＜ III	1				1			
≧ III	2.46	1.537	3.95	**0.00011**	2.342	1.46	3.755	**0.00041**
Gender								
Male	1							
Female	1.359	0.8552	2.159	0.193				
Smoke								
Yes	1							
No	0.8507	0.465	1.556	0.599				
Race								
Asian	1							
Caucasian	0.90824	0.2836	2.909	0.871				
African American	1.80796	0.3646	8.966	0.469				
Hispanic	1.075	0.2567	4.502	0.921				

Supplemental Table 4: Contingency analysis table demonstrating the correlation between DOCK4 expression level, frequency of recurrence, and tumor stage in a cohort of 182 lung ADC patients.

		DOCK4 level	
		Low	High
Recurrence	Yes	26 (28%)	46 (52%)
	No	68 (72%)	42 (48%)

Fisher's exact test: ***p<0.001

		DOCK4 level	
		Low	High
Tumor stage	< III	73 (78%)	56 (64%)
	≥ III	21 (22%)	32 (36%)

Fisher's exact test: *p<0.05

Supplemental Materials and Methods

Cell lines

Lung adenocarcinoma (ADC) cell lines used in this study were: A549, HCC4006, H441, PC9, H1975, and H1793 (gift from R. Sordella, CSHL). All lung ADC cell lines were grown in RPMI 1640, GlutaMax^{TM} (Invitrogen) supplemented with 5% fetal bovine serum (FBS) and 1% penicillin/streptomycin (Invitrogen). The following breast cancer cell lines were used: MDA-MB-134VI, MDA-MB-231, MDA-MB-435S, SK-BR-3, and Hs578T (gift from M. Wigler, CSHL). MDA-MB-134-VI, MDA-MB-231, and MDA-MB-435S were grown in Leibovitz's L-15 medium (ATCC) supplemented with 10% FBS and 1% penicillin/streptomycin. SK-BR-3 cells were grown in McCoy's 5a (ATCC), and Hs578T in DMEM (Invitrogen), each of them supplemented with 10% FBS (Invitrogen) 0.01 mg/ml insulin (Invitrogen), and 1% penicillin/streptomycin. The known somatic mutations for common oncogenic/tumor suppressive pathways in the above cell lines are listed in Supplemental Table 1. For TGF-β treatment, a 1:1 mixture of human recombinant TGF-β1 and TGF-β2 (R&D systems) was added to the culture media at a final concentration of 2 ng/ml.

Plasmids, shRNAs and viral transduction

TtIGP-DOCK4^{WT} and TtIGP-DOCK4$^{\Delta DHR2}$, which lacks an internal deletion of 77 amino acids within the DHR2 domain (residues 1357-1433), were generated by subcloning DOCK4 and DOCK4$^{\Delta DHR2}$ cDNA fragments from pcDNA3-Flag-DOCK4 and pcDNA3-Flag-DOCK4$^{\Delta DHR2}$ (gift from V. Yajnik, MGH) into the retroviral TtIGP-IRES-GFP-puro vector (gift from S.W. Lowe, MSKCC) using *NotI* and *XhoI* restriction sites. MSCV-Rac1^{Q61L} was generated by subcloning Rac1^{Q61L} cDNA from pcDNA3-Rac1^{Q61L} (gift from J. Schlessinger, Yale) into the retroviral MSCV-PGK-GFP-puro vector (gift from S.W. Lowe) using *XhoI* and *EcoRI* restriction sites. The pLenti-III-mCherry-HRE-luciferase (firefly) vector was purchased from Applied Biological Materials. pEF1α-tTA-neo was a gift from P. Kenny (AECOM). For RNAi experiments, DNA fragments encoding two shRNAs directed against the 3' untranslated region (3'UTR) of human *DOCK4* mRNA (Dock4#1: 5'-TTGAGTGTTATAAACAGTCAG-3'; Dock4#2: 5'-TACATTACTACATACAAGGTG-3') were cloned into the retroviral miR-30-based

MLP-PGK-GFP-puro or MLP-PGK-DsRed-neo vectors (gift from S.W. Lowe) using *XhoI* and *EcoRI* restriction sites. The shRNAs targeting human Smad4 (shSmad4#1 and shSmad4#2) were purchased from Addgene (#15724 and #37046). The Rac1 shRNAs (shRac1#1 and shRac1#2) were gifts from Y. Zhang (Cincinnati Children's Hospital). A shRNA targeting Renilla luciferase was used as a negative control and was previously described (Zuber et al. 2011). Cell lines stably expressing cDNAs or shRNAs were generated by retroviral or lentiviral transduction. For the production of viral particles, retroviral vectors were co-transfected with pCL-Ampho and pVSV-G into Phoenix-Ampho cells, and lentiviral vectors were co-transfected with pCMVΔR8.91, pcREV and pVSV-G into HEK-293T cells. The virus-containing medium was collected 48 h after transfection and the target cells were spin infected. Infected cells were selected in media containing 2 μg/ml puromycin for 2 days or 100 μg/ml neomycin for 4 days.

Immunoblotting

For Western blot analyses, cells were lysed in RIPA buffer containing protease (Roche) and phosphatase (Sigma) inhibitors. Proteins were separated by SDS-PAGE and blotted onto PVDF membranes (Bio-Rad). Membranes were probed with the following primary anti-human antibodies: Gapdh (1:5,000, Novartis), DOCK4 (1:2,000, gift from V. Yajnik), phospho-Smad3 (Ser423/425) (1:1,000, Cell Signaling), Smad3 (1:1,000, Cell Signaling), Smad4 (1:1,000, Cell Signaling), E-cadherin (1:10,000, BD Biosciences), phospho-β-catenin (Ser33/Ser37/Thr41) (1:1,000, Cell Signaling), β-catenin (1:5,000, Cell Signaling), Rap1 (1:1000, BD Biosciences), Cdc42 (1:1000, Cytoskeleton), and Rac1 (1:1,000, Cytoskeleton), followed by HRP-conjugated secondary antibodies (1:5,000, Bio-Rad).

Co-immunoprecipitation

pCL-neo-HA-DOCK3 (gift from C. Marshall, ICR) or pcDNA3-Flag-DOCK4 was co-transfected with LZRS-hNEDD9-IRES-GFP, which contains an N-terminal Flag and a C-terminal HA epitope tag (#21962, purchased from Addgene), into HEK-293T cells. 24 h post-transfection, cells were lysed in lysis buffer containing 200 mM Tris pH 8.0, 150 mM NaCl, 0.5% NP-40, with protease and phosphatase inhibitors. 1 mg total protein was incubated with 1 μg anti-Flag M2 (Sigma) or 1 μg anti-HA (Covance) antibody for 1 h at

4°C, followed by incubation with protein A/G Sepharose beads (GE Healthcare) for 1 h at 4°C. Beads were washed 3 times with lysis buffer. Immunoprecipitates were then resolved by SDS-PAGE and immunoblotted with antibodies against HA (1:5,000, Covance), Flag M2 (1:2,000, Sigma), NEDD9 (1:2,000, Abcam), or DOCK4 (1:2,000, gift from V. Yajnik).

Rac1/Cdc42/Rap1-GTP pull-down assay

To assay Rac1, Cdc42, and Rap1 activity, GST-PBD and GST-RalGDS-RBD pull-down assays were performed as described before (Govek et al. 2004; Boettner et al. 2000), using A549 cells expressing indicated constructs.

Immunofluorescence and confocal image acquisition

For immunostaining of cultured lung ADC cells, the cells were grown on coverslips and fixed for 15 min at RT in 4% paraformaldehyde (PFA) in TBS. For immunostaining of liver tissue sections, livers were dissected from mice 20 h after intracardiac injection of lung ADC cells. Animals were deeply anesthetized with isoflurane and perfused transcardially with 4% PFA. Livers were post-fixed in 4% PFA overnight, embedded in 3% agarose, and cut into 50-µm thick sections using a vibratome (Leica VT1000S). Cultured lung ADC cells and liver sections were permeabilized and blocked with 0.1% Triton X-100 and 1% BSA in TBS for 5 min (for cells) or 1 h (for sections) at RT, followed by incubation with primary antibodies diluted in TBS with 1% BSA at 4°C overnight. We used the following primary antibodies: mouse anti-human E-cadherin (1:400, BD Biosciences); rat anti-mouse CD31 (clone MEC13.3, 1:200, BD Pharmingen); and chicken anti-GFP (1:1,000, Aves Labs). The secondary antibodies used were: goat anti-chicken Alexa Fluor 488, goat anti-rat Alexa Flour 594, and goat anti-mouse Alexa Fluor 647 (1:1,000, Molecular Probes). Fluorescence images of lung ADC cells and liver sections were acquired using Zeiss LSM 510 and 780 confocal microscopes with a 63x or 40x oil-immersion objective. In the case of liver sections, images were taken at z-sections (15-30 sections) of 1-µm intervals, and images were reconstructed using Volocity image processing software.

Tissue microarray and immunohistochemistry

Human lung adenocarcinoma tissue microarrays (TMA LC706 and TMA LC10013) were purchased from US Biomax Inc. Immunohistochemical staining with anti-p-Smad3 (#9520, 1:25, Cell Signaling), anti-DOCK4 (#ab56743, 1:400, Abcam), or anti-β-actin (#ab6276, 1:500, Abcam) antibody was performed using DAB chromogen kit (Vector labs) according to the manufacturer's protocol. For quantification of p-Smad3 IHC staining, the Aperio IHC Nuclear Image Analysis algorithm was used. This algorithm detects the staining intensity of nuclei stained with a specific chromagen (brown). Nuclei are also identified by size and shape. The values for the analysis are set by the pathologist (J.E.W). A minimal level for positive staining is set and the nuclear staining is classified as 0, 1+, 2+ and 3+ based on nuclear staining intensity. A nucleus is classified 0 when it has no nuclear staining. 1+ nuclei have weak but positive staining. A nucleus is classified 2+ when it has moderate nuclear staining and 3+ when there is intense staining. Based on the percentages of 0, 1+, 2+ and 3+ nuclei, the final H-score is determined (1 X % weak staining) + (2 X % moderate staining) + (3 X % strong staining). For analysis, 10-20 areas of each core containing only tumor cells were traced. Only these areas were scanned for the analysis. For quantification of DOCK4 IHC staining, the Aperio Cytoplasm Analysis algorithm was used. This algorithm measures the intensity of the stain (brown) in the cytoplasm. The default values for nuclear staining (hematoxylin) and cytoplasmic staining (brown) were used. Cytoplasmic segmentation was included in the analysis and the distance from the nucleus to the cytoplasm was set using a visual check of this value by the pathologist (J.E.W). This parameter defines how far from the nucleus the cytoplasm can be reported as cytoplasm that surrounds the nucleus. For analysis, 10-20 areas containing only tumor or adjacent normal epithelial cells of each core were traced. Only these areas were scanned for the analysis. The intensity of the staining was recorded as 0, 1+, 2+ and 3+ based on cytoplasmic staining intensity. A cytoplasm is classified 0 when it has no cytoplasmic staining. 1+ cytoplasm has weak but positive staining. Cytoplasmic staining is classified 2+ when it has moderate cytoplasmic staining and 3+ when there is intense staining. The thresholds were set by the pathologist based on the most intense staining in the slide and the lowest intensity deemed positive. Based on the percentages of 0, 1+, 2+ and 3+ cytoplasm, the final H-score is determined (1 X %

weak staining) + (2 X % moderate staining) + (3X % strong staining). The H-scores for p-Smad3 and DOCK4 were further converted to a 0-6 scale. The samples with raw H-scores < 5^{th} percentile and > 95^{th} percentile were designated scores of 0 and 6, respectively. In the samples between 5^{th} and 95^{th} percentile, the maximal raw H-score was set as 100%. The samples with raw H-scores between 80-100%, 60-80%, 40-60%, 20-40%, and 0-20% were then designated scores of 5, 4, 3, 2, and 1, respectively. Based on the 0-6 scale, the samples were classified into p-Smad3 or DOCK4 high (score 3-6) or low (score 0-2) groups for contingency analysis. The *P* value was calculated by Fisher's exact test.

RNA isolation and quantitative PCR (qPCR)

Total RNA was isolated from cells using TRIzol reagent (Invitrogen), purified by LiCl precipitation and then reverse transcribed using the TaqMan reverse transcription kit (Invitrogen). The resulting cDNA was amplified by qPCR using Power SYBR Green PCR Master Mix (Invitrogen). qPCR and data collection were performed on an Applied Biosystems (ABI) 7900HT Fast Real-Time PCR system. The primers used in this study are listed in Supplemental Table 2.

Chromatin immunoprecipitation (ChIP)

A549 or MDA-MB-231 cells (~3×10^6) in a 10 cm dish were left untreated or treated with 2 ng/ml TGF-β for 5 h. Chromatin immunoprecipitation (ChIP) assays were performed using a ChIP-IT Express Enzymatic Magnetic Chromatin Immunoprecipitation Kit & Enzymatic Shearing Kit according to the manufacturer's instructions (#53009, Active Motif). The antibodies used in the ChIP assays were normal rabbit IgG (sc-2027, Santa Cruz) and p-Smad3 (Ser423/425) (#9520, Cell Signaling). 4 μg of each antibody and ~50 μg of total chromatin were used in each ChIP reaction. The primers used in ChIP-qPCR analysis are listed in Supplemental Table 2. Enrichment was quantified as the IP/Input ratio using SYBR green reagent on the ABI7900HT.

Cell proliferation and anoikis assays

For 2D cell proliferation assays, 5,000 cells were suspended in regular culture medium and plated into 6-well plates. Cell numbers were counted using a Nexcelom Cellometer

Auto T4 Cell Counter at the indicated time points. For anoikis assays, 10,000 cells were suspended in culture media and plated into 96-well plates pre-coated with poly-HEMA (poly 2-hydroxyethyl methacrylate) to prevent cell attachment. Cells were incubated at 37°C for the indicated time points, and cell viability was measured using a MTT assay (Invitrogen) following the manufacturer's protocol.

Time-lapse microscopy and single cell motility analysis

5,000-10,000 cells were plated into 6 or 12-well plates in culture media supplemented with or without TGF-β for 24 h prior to imaging. Cells were imaged every 6 min over a total period of 4.5 h using a Zeiss Observer automated imaging system. The movement of single cells was tracked and quantified using Zeiss Axiovision 6.0 software. The coordinate graphs were plotted using Origin 8 software; each cell was plotted with time 0 at the origin of the grid. Single cells were randomly chosen for each condition.

Matrigel invasion assay

Invasion assays were performed using 24-well PET Transwell inserts (Costar, 8.0 μm pore size) coated with 20 μl of Matrigel (BD Biosciences) at 37°C for 20 h. 10,000 cells resuspended in culture medium containing 0.2% FBS were plated in the upper chamber of the Transwell insert. Culture medium containing 10% FBS was used as a chemoattractant in the lower chamber. 20 h after plating, cells in the upper chamber were removed with a cotton swab. Cells that had migrated through the filters were fixed in 10% methanol and stained with 0.5% crystal violet. The filters were photographed and cell density was calculated using an Odyssey CLx infrared imaging system.

Animal Studies

All animals were maintained in a specific pathogen-free facility, and all studies were conducted under protocols approved by the CSHL Institutional Animal Care and Use Committee. 5-7 week old NOD/SCID/IL2γ (NSG) mice (National Cancer Institute) were used for all xenografting studies. Lung ADC cells were engineered to stably express firefly luciferase. Prior to all injections, lung ADC cells were harvested at a concentration of $5x10^6$ cells/ml in PBS containing 1 mM EDTA pH 8.0, and in all cases animals were anesthetized with 2% isoflurane before injection. For intracardiac injections, a small skin

incision was made on the left side of the chest, and $5x10^5$ cells were injected into the left ventricle using a 26-gauge needle attached to a 1-ml insulin syringe (BD biosciences). For intrathoracic injections, a small skin incision was made to the left chest wall. A 30-gauge needle attached to a 0.5-ml insulin syringe was inserted directly through the intercostal space into the lung to a depth of about 3-5 mm, and $1x10^5$ cells were injected into the lung parenchyma. In both cases the wounds were closed using surgical wound clips (Fine Science Tool) and wound clips were removed 7 d later. For intrahepatic injections, a 2 cm incision was made in the upper abdomen through the peritoneum. The liver was carefully exposed and $1x10^5$ cells resuspended in 20 µl of a 1:1 mixture of Matrigel and PBS were injected intrahepatically using a 30-gauge needle attached to a 0.3-ml insulin syringe (Terumo). The wound was closed with a 6-0 silk suture (CP Medical Inc.). Animals were monitored until recovery after surgery for at least 30 min. For bioluminescence imaging, animals were injected intraperitoneally with 100 µl of 30 mg/ml D-luciferin (GoldBio) in PBS and anesthetized with 2% isofluorane using a XGI-8 gas anesthesia system (Xenogen). Bioluminescence images were acquired 10-15 minutes after injection using the IVIS-200 Imaging System (Xenogen). For metastasis assays, bioluminescent signals were quantified at the indicated time points and normalized to day 0 from live animals. The mean value of the control group was normalized to 100 for data presentation. Metastases were confirmed by necropsy and histology using *ex vivo* bioluminescent imaging and Hematoxylin & Eosin (H&E) staining. Growth rate in lung and liver was measured as function of photon flux normalized to day 0 in live animals.

Bioinformatics

For gene expression data analysis (Supplemental Fig. 1A), a dataset comprising gene expression profiles of TGF-β treated A549 cells was downloaded from Gene Expression Omnibus (GEO) with accession number GSE17708 (Ranganathan et al. 2007). Probes against all 83 Rho family GEFs were selected. Gene expression levels were normalized to the 0 h time point and then converted to log2 scale. The data were processed and a heat map was generated using Expander 6.0. For DOCK3/4 homology analysis, protein sequences for DOCK3 (accession number: AAP80572.1) and DOCK4 (accession number: AAO73565.1) were downloaded from NCBI. The boundaries of the identified functional

domains were defined based on UniProtKB/Swiss-Prot simulations. Sequence homology was analyzed using VectorNTI 11.0 software.

ChIP-Seq data analysis

Short Read Archives (SRAs) for ChIP-Seq analysis were downloaded from GEO database (GSM1246720 and GSM1246721). After converting SRA to FASTQ format, the sequence reads were mapped to the reference genome assembly NCBI36/hg18 using Bowtie following these criteria: -m1, -v2. To identify ChIP-Seq peaks, we used the MACS version 1.4.0beta (Model based Analysis of ChIP-Seq) peak finding algorithm. A p value threshold of enrichment of 1e-5, and a false discovery rate (FDR) of less than 5% was adjusted to define a list of confident peaks.

Lung adenocarcinoma (ADC) and ER-negative breast cancer clinical data analysis.

For lung ADC clinical association analysis, a publicly available dataset (GSE41271) was downloaded from NCBI Gene Expression Omnibus (GEO). It contains microarray-based gene expression data of 182 human lung adenocarcinomas and clinical follow-up information, including recurrence status, recurrence-free survival intervals, tumor stage, age, gender, smoking history, and race. The probe IDs representing DOCK3 and DOCK4 are ILMN_1723440 and ILMN_1801044, respectively. For ER-negative breast cancer clinical association analysis, a cohort of 269 primary ER-negative breast cancer samples with corresponding recurrence-free survival data was extracted from five publicly available datasets (GSE2034, GSE2603, GSE5327, GSE4922, and GSE7390) downloaded from NCBI GEO. The probe ID representing DOCK4 is 205003_at. To assess the predictive power of DOCK3 and DOCK4 for tumor recurrence, we computed the normalized expression values by Z-scores across all patients (log2 transformation and RMA normalization in package “affy”) and established a univariate Cox proportional hazards model (R-package “survival”). We obtained a smooth estimate of the relationship between DOCK3 or DOCK4 expression levels and recurrence hazard ratio, and defined the optimal cut-offs that yield the highest hazard ratio between low and high expression groups (customized cut-off finder; Budczies et al. 2012). We generated Kaplan-Meier survival curves for patients with low and high DOCK3 or DOCK4 expression using GraphPad Prism. Statistical significance was assessed by introducing the expression

value as a continuous covariate in the Cox model. We applied the univariate Cox proportional hazards model on all available clinical annotations, DOCK3 and DOCK4 expression groups for the lung ADC cohort. Statistically significant covariates from the univariate model were included in a multivariate Cox model (R-package "survival"), which are DOCK4 expression and tumor stage. *P* values were calculated by log-rank test. For the correlation between high DOCK4 expression, frequency of recurrence, and tumor stage, the contingency analyses were performed using GraphPad Prism. *P* values were calculated by Fisher's exact test.

Statistical analysis

The two-tailed Student's *t*-test or Mann-Whitney test was used to compare continuous variables between two groups with parametric or non-parametric distributions, respectively. A Fisher's exact test was used to compare dichotomous variables. A Log-rank test was used for survival analysis. All data are presented as the mean ± SD, unless otherwise noted. Statistical significance was defined as $P < 0.05$ for all tests.

Supplemental References

Boettner B, Govek EE, Cross J, Van Aelst L. 2000. The junctional multidomain protein AF-6 is a binding partner of the Rap1A GTPase and associates with the actin cytoskeletal regulator profilin. *Proc Natl Acad Sci* **97**: 9064-9069.

Budczies J, Klauschen F, Sinn BV, Gyrffy B, Schmitt WD, Darb-Esfahani S, Denkert C. 2012. Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. *PLoS ONE* **7**: 1-7.

Govek EE, Newey SE, Akerman CJ, Cross JR, Van der Veken L, Van Aelst L. 2004. The X-linked mental retardation protein oligophrenin-1 is required for dendritic spine morphogenesis. *Nat Neurosci* **7**: 364-372.

Padua D, Zhang XH, Wang Q, Nadal C, Gerald WL, Gomis RR, Massague J. 2009. TGFβ primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* **133**: 66-77.

Ranganathan P, Agrawal A, Bhushan R, Chavalmane AK, Kalathur RK, Takahashi T, Kondaiah P. 2007. Expression profiling of genes regulated by TGF-beta: differential regulation in normal and tumour cells. *BMC genomics* **8**: 98.

Zuber J, Shi J, Wang E, Rappaport AR, Herrmann H, Sison EA, Magoon D, Qi J, Blatt K, Wunderlich M et al. 2011. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478**: 524-528.